

University of Groningen

Development of cognitive processing and judgments of knowledge in medical students

Cecilio-Fernandes, Dario; Kerdijk, Wouter; Jaarsma, A. D. (Debbie) C.; Tio, Rene A.

Published in:
Medical Teacher

DOI:
[10.3109/0142159X.2016.1170781](https://doi.org/10.3109/0142159X.2016.1170781)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Cecilio-Fernandes, D., Kerdijk, W., Jaarsma, A. D. C., & Tio, R. A. (2016). Development of cognitive processing and judgments of knowledge in medical students: Analysis of progress test results. *Medical Teacher*, 38(11), 1125-1129. <https://doi.org/10.3109/0142159X.2016.1170781>

Copyright

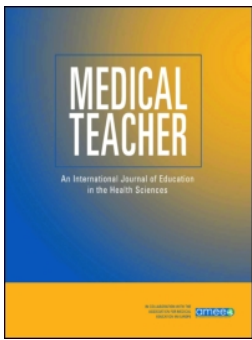
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Development of cognitive processing and judgments of knowledge in medical students: Analysis of progress test results

Dario Cecilio-Fernandes, Wouter Kerdijk, A. D. (Debbie) C. Jaarsma & René A. Tio

To cite this article: Dario Cecilio-Fernandes, Wouter Kerdijk, A. D. (Debbie) C. Jaarsma & René A. Tio (2016) Development of cognitive processing and judgments of knowledge in medical students: Analysis of progress test results, Medical Teacher, 38:11, 1125-1129, DOI: [10.3109/0142159X.2016.1170781](https://doi.org/10.3109/0142159X.2016.1170781)

To link to this article: <http://dx.doi.org/10.3109/0142159X.2016.1170781>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 27 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 448



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

Development of cognitive processing and judgments of knowledge in medical students: Analysis of progress test results

Dario Cecilio-Fernandes^a, Wouter Kerdijk^b, A. D. (Debbie) C. Jaarsma^a and René A. Tio^{a,c}

^aCenter for Education Development and Research in Health Professions (CEDAR), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ^bCenter for Dentistry and Oral Hygiene, Department of Public and Individual Oral Health, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ^cDepartment of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

ABSTRACT

Background: Beside acquiring knowledge, medical students should also develop the ability to apply and reflect on it, requiring higher-order cognitive processing. Ideally, students should have reached higher-order cognitive processing when they enter the clinical program. Whether this is the case, is unknown. We investigated students' cognitive processing, and awareness of their knowledge during medical school.

Methods: Data were gathered from 347 first-year preclinical and 196 first-year clinical students concerning the 2008 and 2011 Dutch progress tests. Questions were classified based upon Bloom's taxonomy: "simple questions" requiring lower and "vignette questions" requiring higher-order cognitive processing. Subsequently, we compared students' performance and awareness of their knowledge in 2008 to that in 2011 for each question type.

Results: Students' performance on each type of question increased as students progressed. Preclinical and first-year clinical students performed better on simple questions than on vignette questions. Third-year clinical students performed better on vignette questions than on simple questions. The accuracy of students' judgment of knowledge decreased over time.

Conclusions: The progress test is a useful tool to assess students' cognitive processing and awareness of their knowledge. At the end of medical school, students achieved higher-order cognitive processing but their awareness of their knowledge had decreased.

Introduction

Students' ability to apply acquired knowledge has been a research topic in medical education for many years (Boshuizen & Schmidt 1992; Eva 2005; Norman 2005). Most studies on knowledge application focus on knowledge growth and differences between beginning and advanced students (for a review, see Wrigley et al. 2012). However, an increase in knowledge does not necessarily imply that students are able to use the acquired knowledge. It may also be achieved through reproduction of factual knowledge, whereas knowledge application requires a deep understanding of factual knowledge. In the course of medical school, students' knowledge becomes more organized, accessible, and hierarchically structured (Bloom 1956; Anderson et al. 2001; Krathwohl 2002), which is also known as students' cognitive processing. Without insight into the cognitive processes involved we are not able to fully help medical students construct hierarchical knowledge.

Bloom's taxonomy is a well-established framework in which cognitive processing is represented as a cumulative hierarchy of lower and higher levels of acquired knowledge. Mastery of the lower levels is required to achieve the higher levels (Bloom 1956; Anderson et al. 2001; Krathwohl 2002). The two lowest levels, remembering and understanding information, are considered lower-order cognitive abilities that require a minimal understanding of information (Crowe et al. 2008). The third level, applying information, is considered a transitional level by some researchers (Crowe et al. 2008), whereas others consider it as a higher-order cognitive

Practice Points

- Preclinical students answer questions using lower-order cognitive processing.
- Clinical students answer questions using higher-order cognitive processing.
- Students' answering patterns correspond with their cognitive abilities.
- Students' judgment of knowledge accuracy decreases over time.
- Progress tests can be used as a tool to measure students' cognitive processing throughout medical school.

ability (Bissell & Lemons 2006). The top three cognitive processes – synthesizing, evaluating, and creating new information – are considered higher-order cognitive skills (Zoller 1993) that require a deep conceptual understanding of the information, but are not necessarily hierarchically structured (Crowe et al. 2008).

Another important aspect of medical students' cognitive processing based upon Bloom's taxonomy is awareness of their own knowledge and cognitive ability. This is known as metacognitive knowledge (Krathwohl 2002). It has been argued that especially medical students should acknowledge what they do not know, because as a doctor they need to make high-stake decisions about patients (Muijtjens et al. 1999). Metacognitive knowledge is usually measured by asking people to provide a judgment of their knowledge about

a specific item. A way to consistently engage students in judging their own knowledge is to add an “I don’t know” option to multiple choice questions. Several studies have investigated incorporation of judgments of knowledge into regular knowledge tests (Keislar 1953; Traub et al. 1969; Muijtjens et al. 1999). These studies generally showed that incorporating an “I don’t know option” increased test reliability and provided valuable information about students’ metacognition. Other studies on self-judgments of knowledge, showed a positive correlation between metacognition and performance (Koriat et al. 2002; Koriat & Shitzer-Reichert 2002; Schleifer & Dull 2009). Furthermore, studies on the effects of experience and study progress on metacognitive knowledge showed that: (1) initial application of knowledge leads to underestimations of one’s own knowledge and (2) metacognitive ability becomes more general rather than domain-specific when students progress through their studies (Koriat & Shitzer-Reichert 2002; Veenman & Spaans 2005). We did not find any studies on the development of undergraduate medical students’ insight into what they do not know and awareness of their knowledge gaps when they progress to more advanced study years.

The most common way of verifying student knowledge is to use tests with different types of questions. First, there are questions that require students to remember and understand basic knowledge. We will refer to these as “simple questions”. Second, there are questions that require students to apply, analyze, and evaluate existing knowledge in combination with new information, which is provided through a case (Crowe et al. 2008). We will refer to these as “vignette questions”. Whereas simple questions aim at assessing lower cognitive processes of Bloom’s taxonomy, vignette questions also require students to use higher cognitive processing which positively affects long-term knowledge retention (Redfield & Rousseau 1981; Jensen et al. 2014).

In this study, we first investigated undergraduate medical students’ cognitive processing by analyzing their answers to simple and vignette questions throughout medical school. We hypothesized that students’ ability to provide correct answers to simple questions would increase because they continuously received theoretical education and had to apply basic, factual knowledge to most of their educational activities. We expected the number of correct answers to vignette questions to increase rapidly when students progressed into the clinical phase, where the emphasis is more on patient cases. We expected the number of incorrect and question mark answers to decrease because student knowledge would increase throughout medical school. Furthermore, we investigated whether students’ self-judgments of knowledge became more accurate over time. We hypothesized that the accuracy of students’ judgments of their own knowledge would increase throughout medical school.

Methods

Study design

We used data from the University of Groningen concerning the Dutch interuniversity progress test of 2008 and 2011 to assess our hypotheses. The progress test is based on the Dutch National Blueprint for the Medical Curriculum and aims to assess the final objectives of undergraduate training, covering the whole domain of medical knowledge at grade.

The Dutch progress test is administered at fixed intervals to all students, four times per year. Each progress test consists of 200 multiple choice questions, comprising simple and vignette questions. Students are allowed to not answer a question by using the “I don’t know” option, hereafter referred to as question mark option. A correct answer is coupled with a reward, an incorrect answer with a penalty and using a question mark ends without reward or penalty (for more details about the Dutch Progress test, see Tio et al. 2016).

From each year, 2008 and 2011, we selected the progress test with the highest reliability, resulting in the first progress test from 2008 ($\alpha = 0.985$) and the last progress test from 2011 ($\alpha = 0.928$). Both tests had similar difficulty levels. For each question we calculated a p value by dividing the number of students who answered the question correctly by the total number of students who answered this question (Crocker & Algina 1986). The overall difficulty of a test is calculated by estimating the mean of all p values within the test. The p values – based on scores from first- to sixth-year medical students from four different medical schools – were 0.34 and 0.37, respectively. Similar p values were found for the University of Groningen: 0.34 and 0.38, respectively.

The sixth-year Groningen undergraduate medical curriculum is divided into a three-year preclinical and a three-year clinical program. As we were interested in students’ cognitive development, we only included data from first-year students from 2008 and last year students from 2011 who participated in one of the two programs. Data of students who did not take both tests were excluded from the dataset.

Data analysis

In accordance with Bloom’s taxonomy, the items of each test were classified as simple or vignette questions by one of the researchers (RT) and a student assistant. Simple questions were items requiring students to remember or/and basically understand the knowledge. Vignette questions were items requiring students to apply, analyze, or/and evaluate existing knowledge. An example of a simple question is:

The blood leaves the liver via the:

- A. Hepatic duct
- B. Hepatic vein
- C. Superior mesenteric vein
- D. Portal vein

An example of a vignette question is:

A 54-year-old male presents with severe head-ache, nausea, and is vomiting since 48 h. At physical examination bilateral papillary edema is present. His blood pressure is 240/160 mmHg. Urinalysis shows proteinuria (2+) and hematuria (1+); no glucose or ketone bodies.

Which of the following nephrologic diseases is most likely?

- A. Acute pyelonephritis
- B. Acute tubulonecrosis
- C. Necrotising arteriolitis
- D. Papillary necrosis

For each test, we determined per student which questions were answered correctly, incorrectly, or with a question mark. As the number of simple and vignette questions

varied between both tests, we calculated percentages for both types of questions.

To analyze students' scores on vignette and simple questions over time, we used a repeated measures analysis of variance (ANOVA) to calculate for each test percentages of correct, incorrect, and question mark answers. For each of the three answering categories, we compared students' first and last year scores on simple and vignette questions. All analyses were separately performed for students in the preclinical and the clinical program.

To assess the accuracy of students' judgments of their own knowledge we calculated a new variable, namely *judgments of knowledge accuracy*. We divided the number of question mark answers by the total number of question mark answers combined with the number of incorrect answers. The formula is as follow:

$$\frac{\text{Question mark answers}}{(\text{Question mark answers} + \text{Incorrect answers})}$$

The underlying assumption was that students fill out a question mark if they do not know the correct answer to a question. In short, the accuracy of students' judgment of knowledge was operationalized as the proportion of answers students did not know out of all the incorrect answers they gave. To compare students' *judgments of knowledge accuracy* between the first and the last year we used paired samples *t*-test. All analyses were separately performed for students in the preclinical and the clinical program.

Results

We used progress test data from 548 first-year preclinical and 411 first-year clinical students. After excluding students who did not take both tests, data from 347 first-year preclinical and 196 first-year clinical students were analyzed.

Percentages of answers are shown in Table 1. As students progressed through their program, the percentage of correct and incorrect answers increased, whereas the percentage of question mark answers decreased.

Preclinical program

For the percentage of correct answers, we found main effects of time ($F(1, 346) = 3800.15, p < 0.001$) and type of question ($F(1, 346) = 76.46, p < 0.001$). Furthermore, we found an interaction effect between year and type of question ($F(1, 346) = 15.48, p < 0.001$). In Year 1, the percentage of correct answers to simple questions was slightly higher than that for vignette questions. In Year 3, the percentage of correct answers to both type of questions increased and the percentage of correct answers to simple questions was higher than that for vignette questions, as compared with Year 1 (Table 1).

For the percentage of incorrect answers, we found main effects of time ($F(1, 346) = 949.69, p < 0.001$) and type of question ($F(1, 346) = 20.03, p < 0.001$). Furthermore, we found an interaction effect between year and type of question ($F(1, 346) = 36.09, p < 0.001$). In Year 1, the percentage of incorrect answers to simple questions was higher than that for vignette questions. In Year 3, the percentage of incorrect answers to both types of questions increased. However, the percentage of incorrect answers to simple

Table 1. Percentage of simple and vignette questions answered correctly, incorrectly, or with a question mark, for the preclinical and clinical program.

| Type of question | Type of answer | Preclinical | | Clinical | |
|--------------------|----------------|-------------|--------|----------|--------|
| | | Year 1 | Year 3 | Year 1 | Year 3 |
| Simple questions | Correct | 8.7 | 44.6 | 36.3 | 55.1 |
| | Incorrect | 7.4 | 25.0 | 17.3 | 26.2 |
| | Question mark | 83.8 | 30.3 | 46.4 | 18.6 |
| Vignette questions | Correct | 7.7 | 42.0 | 36.0 | 61.0 |
| | Incorrect | 5.5 | 25.3 | 19.2 | 25.9 |
| | Question mark | 86.8 | 32.6 | 44.8 | 13.1 |

questions was slightly lower than that for vignette questions, as compared with Year 1 (Table 1).

For the percentage of question mark answers, we found main effects of time ($F(1, 346) = 2746.53, p < 0.001$) and type of question ($F(1, 346) = 135.95, p < 0.001$). However, we did not find an interaction effect between year and type of question ($F(1, 346) = 2.34, p = 0.127$). In Year 3, the percentage of question mark answers was significantly lower than that in Year 1. Furthermore, the percentage of question mark answers to vignette questions was significantly higher than that for simple questions (Table 1).

Clinical program

For the percentage of correct answers, we found main effects of time ($F(1, 195) = 1081.36, p < 0.001$) and type of question ($F(1, 195) = 57.08, p < 0.001$). Furthermore, we found an interaction effect between year and type of question ($F(1, 195) = 89.39, p < 0.001$). We found a similar percentage of correct answers to vignette and simple questions, with the percentage of correct answers to vignette questions being slightly lower. In Year 3, the percentage of correct answers to both type of questions increased. However, the percentage of correct answers to vignette questions was significantly higher than that for simple questions (Table 1).

For the percentage of incorrect answers, we found main effects of time ($F(1, 195) = 145.52, p < 0.001$) and type of question ($F(1, 195) = 5.18, p = 0.024$). Furthermore, we found an interaction effect between year and type of question ($F(1, 195) = 12.99, p < 0.001$). Similar to the preclinical program, the percentage of incorrect answers increased in favor of vignette questions. In Year 4, the percentage of incorrect answers to vignette questions was significantly higher than that for simple questions. In Year 3, the percentage of incorrect answers to vignette questions was slightly lower than that for simple questions (Table 1).

For the percentage of question mark answers, we found main effects of time ($F(1, 195) = 734.91, p < 0.001$) and type of question ($F(1, 195) = 76.90, p < 0.001$). Furthermore, we found an interaction effect between year and type of question ($F(1, 195) = 35.05, p < 0.001$). Similar to the preclinical program, the percentage of question mark answers decreased. In this case, the percentage of question mark answers to simple and vignette questions was similar in both years; however, the decrease in question mark answers to vignette questions was larger than that for simple questions (Table 1).

Judgments of knowledge accuracy

Table 2 shows the outcomes of the paired sample *t*-test comparison of students' *judgments of knowledge accuracy*

Table 2. Mean, paired sample *t*-test and significance values of the variable *judgments of knowledge accuracy* for simple, vignette, and total number of questions in the preclinical and clinical program.

| Type of questions | Preclinical | | | Clinical | | |
|-------------------|-------------|--------|--------------------|----------|--------|--------------------|
| | Year 1 | Year 3 | <i>t</i> | Year 1 | Year 3 | <i>t</i> |
| Simple | 0.91 | 0.52 | 29.53 ^a | 0.71 | 0.39 | 18.92 ^a |
| Vignette | 0.94 | 0.53 | 29.36 ^a | 0.69 | 0.31 | 20.82 ^a |
| Total | 0.92 | 0.52 | 29.87 ^a | 0.71 | 0.37 | 19.92 ^a |

^aThe significant value is corrected by Bonferroni $\alpha = 0.05/3 = 0.017$.

between the first and the last year of the preclinical and the clinical program. In both programs, students' judgments of knowledge accuracy decreased as students progressed through the preclinical and clinical programs.

Discussion

In this study we hypothesized that, due to increasing cognitive processing, students' ability to provide more correct answers to simple and vignette questions would increase. In line with this hypothesis, we found that the percentage of correct answers to both types of questions increased as students progressed through the curriculum. In the preclinical years and the first year of the clinical programme, the percentage of correct answers to simple questions was higher compared with vignette questions. However, at the end of the curriculum the percentage of correct answers to vignette questions was higher compared with simple questions. This confirms our second hypothesis that clinical experience can help students identify correct answers to vignette questions. Our findings may imply, therefore, that students increasingly engage in higher levels of cognitive processing throughout medical school. Additionally, we expected students' self-judgment of knowledge to become more accurate over time. However, we found a decrease in students' *judgments of knowledge accuracy*. As students progressed through both the preclinical and clinical program they provided more correct but also more incorrect answers to progress test questions.

The observed decrease in students' *judgments of knowledge accuracy* is not in line with the literature on metacognition, stating that subjects with higher knowledge levels have higher metacognition ability than subjects with lower knowledge levels (Maki, Jonas & Kallod 1994; Kruger & Dunning 1999). Students in later years seemed to underestimate their knowledge compared with novice students (Kampmeyer et al. 2015). One explanation may be that students may have weighed the probability and degree of benefit of a correct answer against the probability and degree of penalty of an incorrect answer. The outcome of this weighing process has been shown to depend heavily on the penalty of an incorrect answer (Espinosa & Gardezabal 2010). If a penalty was not considered to be sufficiently high, risk-taking behaviors may have been increased during the test. Another explanation for finding a decrease in students' *judgments of knowledge accuracy* concerns the use of the progress test as an assessment tool. As students are expected to score higher in subsequent years, their strategies to answer questions might have changed as well. Alternatively, students might have become overconfident about their knowledge due to experience. It has been shown that clinical encounters and participation in clinical practice builds students' self-confidence (Harrell et al. 1993; Cleave-Hogg & Morgan 2002; Dornan et al. 2005). However,

self-confidence is not necessarily predictive of performance (Harrell et al. 1993; Cleave-Hogg & Morgan 2002). It might have been further enforced by hindsight bias, referring to health care situations where people overestimate the extent to which they would have known something that just happened (Arkes et al. 1981).

Strengths and limitations

A distinctive feature of this study is the use of students' progress test results, which eliminates bias regarding willingness to participate in our study. Although the progress test is a valid and reliable assessment tool for measuring factual knowledge (Muijtjens et al. 1999; Schuwirth & van der Vleuten 2012; Wrigley et al. 2012), we demonstrated that the progress test can also be used to assess students' cognitive processing and the accuracy of their self-judgments of knowledge.

Due to a limited number of places in the clinical program at the time of our study, students were enrolled at different times. Therefore, we were not able to use the same sample of students and we had to analyze the data of both programs separately. Another limitation of our study may be that we used data from a single university. The outcomes may differ from those of other universities with different curricula. However, the underlying cognitive development should be similar at student level, which means that our findings should be replicable across universities and curricula. As guessing is heavily influenced by risk-taking behavior, the use of formula scoring might produce bias regarding students' answers. For example, male students tend to guess more often than female students (Budescu & Bar-Hillel 1993). One might argue that the formula scoring may have blurred the findings of our study. However, students' awareness of their knowledge is part of the cognitive system and by giving them an option to not answer the questions we force them to reflect on their knowledge. Research on self-regulation revealed that students are more able to assess whether they can answer specific questions than to perform a self-assessment (Eva & Regehr 2007, 2011). However, our findings demonstrated that students in later years, who were sitting a high-stakes assessment, rather answered questions they did not know the answer to.

In a more general sense, the retrospective character of this study does not allow us to control for many other variables that may have influenced its outcome. However, laboratory research, which allows to control all variables, has been criticized due to the lack of reproducibility in real life situations. Within the educational environment, the Dutch progress test offers a unique opportunity to study students' cognitive processing and judgments of knowledge in a naturalistic setting.

Practical implications and future research

It may be beneficial for student knowledge acquisition when the learning environment is tailored to students' current state of cognitive processing. Students may not be able to identify their own knowledge gaps in the last year of medical school, which may – in extreme cases – cause possible harm to patients. Furthermore, our study revealed that whether students will answer a progress test question may not be related to judgment of knowledge or self-

regulation, because students in later years may have adapted their answering strategies.

Future research should explore and increase the understanding of cognitive aspects of curriculum design. Additionally, further studies are necessary to better understand why students do not answer questions that require higher-order cognitive processing earlier in their medical training. Finally, if self-judgment of knowledge is a desired feature for progress tests, further research should determine the optimal penalty for incorrect answers.

Conclusions

Preclinical students reproduced their knowledge through lower-order cognitive processing, whereas clinical students applied their knowledge through higher-order cognitive processing. The accuracy of students' judgments of knowledge decreased over time.

Acknowledgements

The authors would like to thank Mrs Tineke Bouwkamp-Timmer for her feedback on the final version of the article and editorial help.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Funding information

This work was supported by CAPES – Brazilian Federal Agency for Support and Evaluation of Graduate Education – [grant 9568-13-1] awarded to D.C.-F.

Notes on contributors

Dario Cecilio-Fernandes is a PhD student, Center for Education Development and Research in Health Professions (CEDAR), University of Groningen and University Medical Center Groningen.

Wouter Kerdijk, PhD, is an administrator, teacher, and educational researcher at the Center for Dentistry and Oral Hygiene, Department of Public and Individual Oral Health, University of Groningen and University Medical Center Groningen.

A. D. (Debbie) C. Jaarsma, DVM, PhD, is Professor of Innovation & Research in Medical Education and chair of the Center for Education Development and Research in Health Professions (CEDAR) at the University of Groningen and University Medical Center Groningen.

René A. Tio, MD, PhD, is an associate professor, Center for Education Development and Research in Health Professions (CEDAR) and Department of Cardiology, University of Groningen and University Medical Center Groningen.

References

- Anderson LW, Krathwohl DR, Bloom BS. 2001. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman.
- Arkes HR, Wortmann RL, Saville PD, Harkness AR. 1981. Hindsight bias among physicians weighing the likelihood of diagnoses. *J Applied Psychol.* 66:252.
- Bissell AN, Lemons PP. 2006. A new method for assessing critical thinking in the classroom. *BioScience.* 56:66–72.
- Bloom BS. 1956. Taxonomy of educational objectives: the classification of education goals. Cognitive domain. Handbook 1. New York: Longman.
- Boshuizen HPA, Schmidt HG. 1992. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cog Sci.* 16:153–184.
- Budescu D, Bar-Hillel M. 1993. To guess or not to guess: a decision-theoretic view of formula scoring. *J Educ Meas.* 30:277–291.
- Cleave-Hogg D, Morgan PJ. 2002. Experiential learning in an anaesthesia simulation centre: analysis of students' comments. *Med Teach.* 24:23–26.
- Crocker LM, Algina J. 1986. Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.
- Crowe A, Dirks C, Wenderoth MP. 2008. Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ.* 7:368–381.
- Dornan T, Scherpbier A, King N, Boshuizen H. 2005. Clinical teachers and problem-based learning: a phenomenological study. *Med Educ.* 39:163–170.
- Espinosa MP, Gardeazabal J. 2010. Optimal correction for guessing in multiple-choice tests. *J Math Psychol.* 54:415–425.
- Eva KW. 2005. What every teacher needs to know about clinical reasoning. *Med Educ.* 39:98–106.
- Eva KW, Regehr G. 2007. Knowing when to look it up: a new conception of self-assessment ability. *Acad Med.* 82(Suppl. 10):S81–S84.
- Eva KW, Regehr G. 2011. Exploring the divergence between self-assessment and self-monitoring. *Adv Health Sci Educ Theor Pract.* 16:311–329.
- Harrell P, Kears G, Reed E, Grigsby D, Caudill T. 1993. Medical students' confidence and the characteristics of their clinical experiences in a primary care clerkship. *Acad Med.* 68:577–579.
- Jensen JL, McDaniel MA, Woodard SM, Kummer TA. 2014. Teaching to the test...or testing to teach: exams requiring higher order thinking skills encourage greater conceptual understanding. *Educ Psychol Rev.* 26:307–329.
- Kampmeyer D, Matthes J, Herzig S. 2015. Lucky guess or knowledge: a cross-sectional study using the Bland and Altman analysis to compare confidence-based testing of pharmacological knowledge in 3rd and 5th year medical students. *Adv Health Sci Educ.* 20:431–440.
- Keislar ER. 1953. Test instructions and scoring method in true-false tests. *J Exp Educ.* 21:243–249.
- Koriat A, Sheffer L, Ma'ayan H. 2002. Comparing objective and subjective learning curves: judgments of learning exhibit increased under-confidence with practice. *J Exp Psychol Gen.* 131:147.
- Koriat A, Shitzer-Reichert R. 2002. Metacognitive judgments and their accuracy. New York, NY: Springer.
- Krathwohl DR. 2002. A revision of Bloom's taxonomy: an overview. *Theor Pract.* 41:212–218.
- Kruger J, Dunning D. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 77:1121.
- Maki RH, Jonas D, Kallod M. 1994. The relationship between comprehension and metacomprehension ability. *Psychol Bull Rev.* 1:126–129.
- Muijtjens AMM, Mameren HV, Hoogenboom E, van der Vleuten CPM. 1999. The effect of a 'don't know' option on test scores: number-right and formula scoring compared. *Med Educ.* 33:267–275.
- Norman G. 2005. Research in clinical reasoning: past history and current trends. *Med Educ.* 39:418–427.
- Redfield DL, Rousseau EW. 1981. A meta-analysis of experimental research on teacher questioning behavior. *Rev Educ Res.* 51:237–245.
- Schleifer LL, Dull RB. 2009. Metacognition and performance in the accounting classroom. *Issues Account Educ.* 24:339–367.
- Schuwirth LWT, van der Vleuten CPM. 2012. The use of progress testing. *Perspect Med Educ.* 1:24–30.
- Traub RE, Hambleton RK, Singh B. 1969. Effects of promised reward and threatened penalty on performance of a multiple-choice vocabulary test. *Educ Psychol Meas.* 29:847–861.
- Tio RA, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJA. 2016. The progress test of medicine: the Dutch experience. *Perspect Med Educ.* 5:51–55.
- Veenman MV, Spaans MA. 2005. Relation between intellectual and metacognitive skills: age and task differences. *Learn Individ Differ.* 15:159–176.
- Wrigley W, van der Vleuten CPM, Freeman A, Muijtjens A. 2012. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach.* 31:683–697.
- Zoller U. 1993. Are lecture and learning compatible? Maybe for LOCS: unlikely for HOCS. *J Chem Educ.* 70:195.